A robust Expectation-Maximization method for the interpretation of small angle scattering data on dense nanoparticle samples

M. Bakry,[a]* H. Haddar[a] and O. Bunău[b]

[a]*INRIA-Saclay, 1 rue Honoré d'Estienne d'Orves Bâtiment Alan Turing Campus de l'École Polytechnique 91120 Palaiseau France, and* [b]*Xenocs SAS - Headquarters 19 rue François Blumet 38360 Sassenage France. E-mail: marc.bakry@gmail.com*

Abstract

The Local Monodisperse Approximation (LMA) is a two-parameter model commonly employed for the retrieval of size distributions from the small angle scattering (SAS) patterns obtained on dense nanoparticle samples (e.g. dry powders and concentrated solutions). This work features a novel implementation of the LMA model resolution for the inverse scattering problem. Our method is based on the *Expectation Maximization* iterative algorithm and is free from any fine tuning of model parameters. The application of our method on SAS data acquired in laboratory conditions on dense nanoparticle samples is shown to provide good results.

1. Introduction

The design of tools and methodology for the characterization of nanoparticles (NP), in particular their sizing, is a major challenge in the fields of industrial preparation

of nanomaterials (Potocnik, 2011) and environmental regulations (Rasmussen *et al.*, 2016). In this context small angle scattering (SAS) is gaining grounds against the more commonly used electron microscopy techniques, due to minimal sample preparation requirements and statistical relevance of the result.

The interpretation of SAS data in terms of NP sizing is a challenging task. For diluted NP samples the typical methods relevant for the size determination with no hypothesis on the form of the size distributions are based on the Monte-Carlo approach (Bressler *et al.*, 2015; Pauw *et al.*, 2013), variational techniques (Glatter, 1977; Brunner-Popela & Glatter, 1997; Weyerich *et al.*, 1999) or, more recently, the *Expectation Maximization* (EM) optimization scheme (Benvenuto *et al.*, 2016). However these methods fail to provide reliable sizing from SAS data measured on dense NP samples, where concentration effects are significant. As a workaround, dense samples are either diluted prior to the SAS measurement to facilitate data interpretation or measured as such and modelled as non-interacting NP on a restricted data range where the effects of particle correlations are supposedly negligible (Rieker *et al.*, 1999; Brunner-Popela *et al.*, 1999). There exist few interpretations of SAS data on dense samples that describe the entire data range with an interacting model (Pedersen, 1994; Ehmann *et al.*, 2013; Brunner-Popela *et al.*, 1999; Kremser *et al.*, 2012; Bressler *et al.*, 2015) and amongst these only a notable minority deals with dry powders (Kremser *et al.*, 2012; Bressler *et al.*, 2015). All the analysis cited above require user adjustable values such as smoothness constrains or the setting of model parameters, which adds complexity to the SAS data interpretation in concentrated systems.

Accurate NP sizing from SAS data involves the design of a robust computational tool that is applicable irrespective of the sample concentration, makes no hypothesis on the form of the size distribution and ideally has no adjustable parameters, neither for describing the particle interactions nor in the form of smoothness constrains. Nev-

ertheless a generic shape must be assumed for the NP, since the distribution of sizes and the particle form factor cannot be simultaneously extracted from the same data set (Feigin & Svergun, 1987). This work features such a calculation tool all by focusing on spherical NP with well defined interfaces (hard spheres).

The interaction models that have been proposed in the literature are based on a statistical description of particle interactions (Torquato, 2002), which depend on the dielectric properties of the NP and on their sizes. In the following we will focus on hard sphere potentials within the Percus-Yevick closure relation (HSPY) (Vrij, 1978; Vrij, 1979). In particular, the local monodisperse approximation (LMA) is a simplifying limit of HSPY where interactions are restricted in-between particles of identical sizes (Kinning & Thomas, 1984). LMA amounts to a two-parameter linear model (Pedersen, 1997), distinct from the one describing the dilute limit, and is generally contained in software packages dedicated to treating SAS data (Bressler *et al.*, 2015; Breßler *et al.*, 2015; Pedersen, 1997). LMA is relatively easy to implement and solve, due to the HSPY (and subsequently LMA) having an exact, analytical solution (contrary to other interaction potentials and / or closure relations) and to the linear nature of the model.

The main focus of this work is to feature a robust and highly efficient numerical implementation of the LMA model with an EM algorithm. Contrary to alternative implementations of LMA, ours is parameter-free in the sense that an automatic search is performed in order to find the best model parameters with respect to the fitted data. As our method does not involve any regularization scheme, it is free from optimization-related (e.g. smoothness) parameters. The accuracy of the fit is measured using the more complete HSPY model. This automatic parameter optimization is feasible due to the low computational cost of EM as compared to Monte-Carlo or variational methods.

This article is organized as follows. Section 2 contains a synthetic view of the LMA

model as well as its formulation in terms of the EM framework. Section 3 reports the algorithmic details on its resolution and proves the validity of the method by applying it to simulated data. Section 4 contains the application of the method to experimental data issued on three distinct samples and the discussion of the results. The concluding remarks are featured in section 5 whereas the appendix contains the formal description of the EM minimization scheme as well as the various models implemented within this work.

2. Theory

2.1. The LMA model

For an isotropic system of interacting particles the normalized scattered intensity (i.e. the scattering probability *per* unit of sample thickness) reads (see for example (Egami & Billinge, 2003), 3.1.5 on page 65 for more details) :

$$I_{\mathrm{s}}(Q) = \Delta\rho_{SL}^2 \sum_{k,l=1}^{n} \sqrt{V(R_k)V(R_l)} F_{R_k}^*(Q) F_{R_l}(Q) \sqrt{\rho_k \rho_l}\, S_{kl}(Q) \tag{1}$$

where

$$\rho_k = \frac{n_k V_k}{V} \tag{2}$$

is the volume fraction of the k size bin with n_k the number of particles in the species, V the illuminated volume and Q is the modulus of the scattering vector. $F_{R_k}(Q)$ is the form factor associated to the particles of size R_k and volume V_k and $\Delta\rho_{SL}$ is the apparent scattering length density. In eq. (1) the partial structure factor S_{kl} is a dimensionless quantity describing the interaction between the particles of radius R_k and the particles of radius R_l (refer to 6.3 for details).

For non-interacting particles the partial structure factor reduces to the identity matrix $S_{kl} = \delta_{kl}$ and eq. (1) simplifies to :

$$I_{\mathrm{s}}(Q) = \Delta\rho_{SL}^2 \sum_{k=1}^{n} V(R_k) |F_{R_k}(Q)|^2 \rho_k \tag{3}$$

For dense NP systems S_{kl} is a matrix whose elements depend on the size distribution $\{\rho_k\}$ of the sample. Thus (1) is in general a strong non-linear problem which admits an analytical solution in some particular cases such as the HSPY model (see 6.3).

LMA stands for the local monodisperse approximation to the HSPY model (see 6.3) and consists in assuming that particles of a given size are assumed to be surrounded by, and interacting with, other particles of the same size only ($S_{kl} = S_{kk} \cdot \delta_{kl}$). Following the formulation in reference (Pedersen, 1997), the interactions are reduced to those of a single population k parametrized with the hard-sphere volume fraction $\rho^\star \equiv \rho_k$ and its hard-sphere radius $R^\star = C^\star \cdot R_k$. The associated scattered intensity reads (Pedersen, 1997):

$$I_{\mathrm{LMA}}(Q) = \Delta\rho_{SL}^2 \sum_{k=1}^{n} V(R_k) \cdot |\mathcal{F}_{R_k}(Q)|^2 \cdot S(Q, \rho^\star, C^\star \cdot R_k) \cdot \rho_k. \tag{4}$$

The LMA model is therefore a two-parameter $\{C^\star, \rho^\star\}$ model, linear in the unknowns ρ_k. In the ideal monodisperse case, $\rho^\star$ corresponds to the actual volume fraction of the particles and $C^\star$ has a value close to 1 meaning that the interaction radius is similar to the particles radius. In the previous LMA implementations reported in the literature these parameters are either set in advance (Bressler *et al.*, 2015) or optimized through a least-square method (Pedersen, 1994).

2.2. The EM solution to the LMA model

EM (formally described in 6.1), was first implemented in the context of interpreting SAS data on diluted NP samples in reference (Benvenuto *et al.*, 2016). The diluted NP problem (3) is recast into a linear form as

$$I = [\mathbf{H}(Q, R)] \cdot \rho \tag{5}$$

with $I = (I_i)_{i \in [\![1,m]\!]}$ the experimental data assuming a discrete sampling $Q = (Q_i)_{i=1,m}$ and $\rho = (\rho_j)_{j \in [\![1,n]\!]}$ the vector of unknowns (distribution of sizes). $\mathbf{H}(Q, R)$ is the matrix

with entries

$$\mathbf{H}_{ij} = V(R_j)|F_{R_j}(Q_i)|^2\Big|_{(i,j)\in[\![1,m]\!]\times[\![1,n]\!]}.$$

The implementation featured in this work exploits the linearity of the LMA model, which makes it a good candidate to be solved within the EM framework. Indeed the corresponding inverse problem (4) can be recast to :

$$I = [\mathbf{H}(Q,R) \odot \mathbf{S}(Q,R;C^\star,\rho^\star)] \cdot \rho \tag{6}$$

where $\odot$ indicates the element-wise matrix multiplication. The partial factor $S(Q,R;C^\star,\rho^\star)$ is implemented according to the form in reference (Kinning & Thomas, 1984) (see detail in 6.2 for the formulas related to the computation of the LMA model). For a given pair $\{C^\star,\rho^\star\}$ the inverse problem is linear as the interaction term $S(Q,R)$ does not depend on the unknown ρ, whose matrix elements are defined in (2).

Note that, while the inverse problem (6) could be reformulated for the unknowns number concentrations n_k/V instead of volume fractions $\rho_k = n_k V_k/V$, the former is fundamentally more numerically unstable than the latter. In other words the $\mathbf{H} \odot \mathbf{S}$ matrix describing the scattering problem in terms of number concentration has a smaller conditioning number than the one of the matrix describing the problem where the unknowns are the particle volume fractions. As such, the former is more sensitive to experimental noise as compared to the latter and its solution is often unreliable. For this reason we chose to determine size distributions as volume fractions all through this work.

Similar to the proof in 6.1 it can be formally shown that the EM algorithm applied to the LMA problem does not suffer from local minima effects. In other words a unique solution exists and the algorithm is guaranteed to find it for any $\{C^\star,\rho^\star\}$ pair.

3. Implementation and proof of concept

Our approach consists in looking for a solution that is acceptable from the HSPY point-of-view but was obtained within LMA. In other words we use the LMA model to solve the scattering problem while the more complete HSPY model will be used to control the accuracy of the solution. If the data fit is satisfactory (experimental data *versus* HSPY), the solution is validated.

In order to solve the scattering problem (6), we aim at determining the best LMA model parameters $\{C^\star, \rho^\star\}$ in terms of description of the experimental data. We perform a brute-force exploration of the parameters on a linear grid of values. Default search ranges are $\rho^\star \in [0, 1]$ and $C^\star \in [0.8, 1.2]$ with a typical step of 0.01. Search ranges can be restrained if some knowledge on the sample is available. For instance we typically use $\rho^\star \in [0, 0.4]$ and $\rho^\star \in [0.4, 0.8]$ for solutions and powders, respectively. More selective ranges may be used to save computation time. If the minimum is not well defined one should decrease the search step. If the minimum is found next to one of the bounds, the interval should be shifted.

For each couple $\{C^\star, \rho^\star\}$, the iterative, parameter - free *Expectation Maximization* (*EM*, see Appendix 6.1) algorithm is applied to determine the size distribution ρ most likely (i.e. in the sense of maximizing the likelihood) to describe the SAS data $I(Q)$ according to the LMA model in (6). However, the effective reconstruction accuracy will be measured by feeding the *EM*-solution (the model scattering corresponding to the size distribution ρ) to the more complete HSPY model (see Appendix 6.3). The full procedure of obtaining the HSPY model intensity from the LMA solution is described in Appendix 6.3. The best LMA model parameters are chosen by minimizing the χ^2 cost functional with respect to the HSPY model:

$$\chi^2(I_{\text{exp}}, I_{\text{HSPY}}, \rho) = \frac{1}{m} \sum_{i=1}^{m} \frac{\left| I_{\text{exp},i} - I_{\text{HSPY},i}(\rho) \right|^2}{\sigma_i^2} \tag{7}$$

where σ_i is the uncertainty on the i-th measured intensity value $I_{\mathrm{exp},i}$ and $I_{\mathrm{HSPY},i}(\rho)$ is the corresponding model-generated intensity at Q_i. This function allows one to define an easy criterion for assessing the reconstruction quality: $\chi^2 \leq 1$ means that on average, the computed solution lies within the data uncertainty.

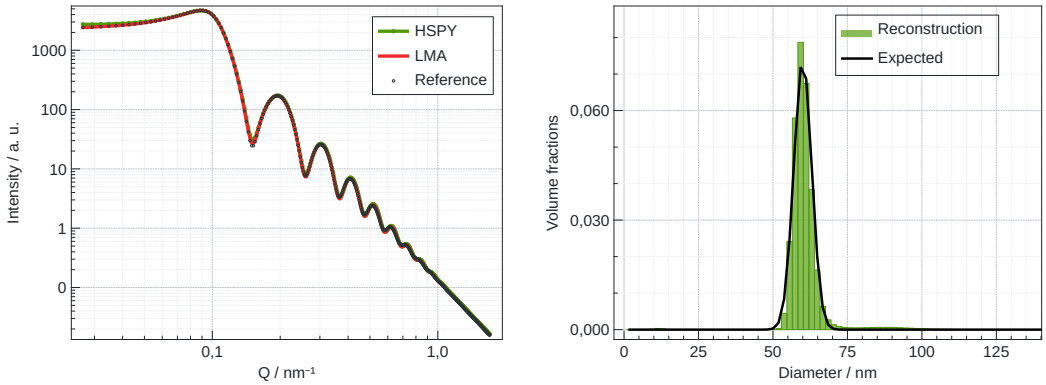
Note that we have equally tried an algorithm based solely on the LMA model for both solving the scattering problem and choosing the best couple of model parameters. While more efficient in terms of computation time, it often gave un-physical solutions for the size distributions in spite of the excellent data fits. This motivated us to refine the values of the $\{C^\star, \rho^\star\}$ parameters based the χ^2 distance between the measure and the HSPY model intensity (instead of LMA).

Note that the model parameter search on a grid is computationally possible thanks to the intrinsic CPU efficiency of the EM algorithm, as a single run of the algorithm may take less than a second. We preferred this brute force method to other optimization techniques such as the gradient descent or the least-squares method (Pedersen, 1994) to ensure that we explore the entire range of parameters. This procedure guarantees the optimality of the particular LMA solution - indeed, we have noticed the existence of several local minima in most of the cases, indicating that the least-squares or gradient descent optimizations are inappropriate for solving this problem.

However, there is no guarantee of success : if, after having explored the entire space of model parameters $\{C^\star, \rho^\star\}$, no reasonable fit was found, it implies the LMA model is not a good enough approximation in order to describe the interactions in the sample. Nonetheless, a good data fit with respect to both the full HSPY model and its simplifying limit LMA is a strong factor in favor of the validation of the solution.

The vector of sizes R of the LMA problem (6) contains linearly spaced values with step $\Delta R = 0.5 \cdot \pi / Q_{\max}$ and covers values up to $R_{\max} = 0.5 \cdot \pi / \Delta Q$ with $Q_{\max}$ the highest scattering vector measured and ΔQ the data resolution.

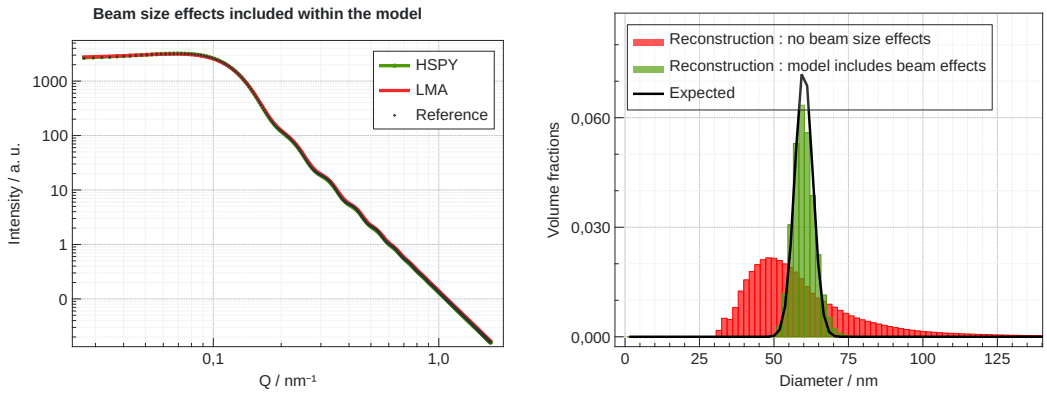
We applied our method on simulated scattering data in order to prove its accuracy. The reference curve was obtained by calculating the HSPY scattering of a 60 nm Gaussian distribution with a total volume fraction of 0.3, then EM-LMA was applied to the reference data (see Fig. 1). EM-LMA succeeds in retrieving the original sizes used to generate the data. We obtain an excellent fit with both LMA and HSPY models, indicating that LMA is a good approximation for HSPY in this particular case. Note that this may no longer be the case for powder-like volume fractions (around 0.6) or very polydisperse distributions.



Size reconstruction on simulated data corresponding to the scattering of 60 nm NP of total volume fraction 0.3 and obeying a Gaussian distribution: fit (left) and distribution of volume fractions (right). EM-LMA succeeds to retrieve the size distribution used to generate the reference scattering curve.

The influence of the finite size of the incident beam is taken into account within the forward model in the form of a convolution kernel applied to the model $\mathbf{H}(Q, R)$. Failing which, the reconstruction misinterprets the experimental smearing and thus results are biased with artefacts, i.e. will contain size populations that are not actually real. We acknowledge it is far more reliable to describe smearing within the model and invert the raw data, than to desmear the data and apply the no-smearing model upon it. The latter is more prone to artifacts than the former as desmearing noised data is yet another source of uncertainty.

The importance of including the beam size effects within the model is illustrated on simulated data, as follows. To emulate the finite beam effects we convoluted the reference scattering curve described above with a Gaussian of width 0.075 nm^{-1} at half maximum, standing for the point spread function (PSF) associated to the instrumental resolution. EM-LMA is applied on the convoluted data, with and without taking into account the convolution kernel in the forward model, as described in the previous paragraph. The results are featured in Fig. 2. Size reconstruction fails if the beam size effects are not described within the model, while the initial size distribution is retrieved upon inclusion of the convolution kernel. Generally speaking the inclusion of the beam size effects become paramount when the width of the PSF is of the same magnitude as the oscillations in the scattering curve.



Size reconstruction on simulated data corresponding to the scattering of 60 nm NP for an incident beam of finite size. On the left, the expected intensity is satisfactorily fit upon inclusion of finite beam size effects in the forward model. On the right we compare the reconstructed volume fractions with and without considering this effect. It clearly appears that inclusion on beam resolution effects is essential for retrieving the sizes.

A Monte Carlo procedure was implemented on top of EM-LMA in order to estimate the uncertainties on the calculated volume fractions. This procedure assesses the numerical stability of the reconstruction and is indirectly correlated with the noise level in the data, i.e. the noisier the data, the higher the uncertainties of the results. A first

iteration is run in order to refine the model parameters and get the LMA solution, as described previously. Once the solution is found (i.e. for a given pair of model parameters) we test its stability by applying a random Poisson noise on the corresponding model intensity and subsequently invert the newly obtained curve within EM-LMA. The process is repeated 100 times. The uncertainties are calculated by taking the standard deviation of the solutions obtained during the Monte Carlo cycles. Typically we notice that about 10 cycles are enough for uncertainties to converge with respect to the number of Monte Carlo cycles.

4. Results and discussion

In this section we apply the previously introduced method to small angle X-ray scattering (SAXS) data acquired on a laboratory instrument on three distinct samples (two powders and a concentrated dispersion) and extract the NP size histogram. The measurements were performed with commercial laboratory SAXS devices (Xeuss and Nano-inXider) with a bidimensional hybrid pixel detector at Xenocs headquarters. Data was corrected similarly but not identically to the procedure described in (Pauw *et al.*, 2017). In particular, the raw 2D data was corrected for invalid pixels, pixel dead time and detector flat-field. The corrected 2D data was azimuthally integrated and corrected for geometrical effects then normalized to the number of transmitted photons (flux transmitted by the sample times exposure time) and to the solid angle seen by the pixel intersecting the direct beam. Poisson statistics is assumed for the corrected 2D data and data uncertainties are propagated accordingly throughout the integration and normalization steps. Finally the background contribution (i.e. scattering of the sample-holder in the case of powders) is subtracted from the 1D (integrated) data with propagation of uncertainties.

In spite of the above corrections, subtracted data systematically features a constant

background term B whose origin is not NP scattering related. This contribution is estimated by fitting the experimental data in the high Q range with a Porod law model function $AQ^{-4} + B$. The constant B is subtracted from the data prior to calculations.

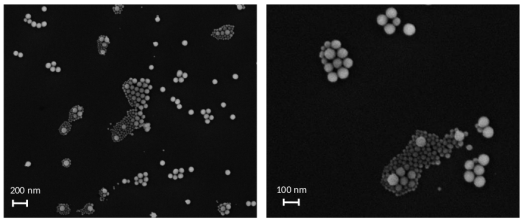
4.1. Sample 1: NP bipopulation in concentrated solution

Sample 1 is a concentrated (nominal 40% volume fraction) dispersion of bi-modal SiO_2 NP in aqueous solution, available commercially (Nyacol, 2017) (NexSil 85-40). Sample 1 was measured for a recent round-robin metrology campaign aiming to compare particle sizing results obtained by multiple techniques, including SAXS, and on various instruments (Feltin, 2018*a*). As such, Sample 1 (labeled "2" in reference (Feltin, 2018*a*)) has already been thoroughly characterized and is known to contain a main population at 70 nm in diameter and a secondary, lower size mode at 30 nm (Feltin, 2018*b*) (see Fig. 3). The SAXS data acquisition was performed on the Xeuss 2.0 device with a Pilatus 300k detector (Dectris) situated at 2609 mm from the sample, for a 30 minute exposure time. The buffer was unavailable and therefore its contribution was not subtracted from the data. While this adds as a source of uncertainty and prevents one from having absolute units, it does not prevent the data treatment. The signature of an aqueous buffer is almost a constant for most of the considered Q range, except for low values where its contribution is negligible whatsoever, given the concentration of the sample. The contribution of the aqueous buffer is partially absorbed in the background constant B and therefore corrected for, as such. The lack of absolute units is dealt with according to the procedure described in 6.3.

The solution obtained with our LMA implementation and the corresponding data fits with respect to both models is represented in Fig. 4. The two modes are identified at 74 and 44 nm, respectively, where the mean values have been estimated as $\sum_k \rho_k \cdot R_k / \sum_k \rho_k$ on the corresponding ranges. The size reconstruction finds the main

population at the expected position while it slightly overestimates the size of the secondary mode. The optimized value of the LMA parameter $\rho^\star = 0.38$ is consistent with the nominal value of the total volume fraction.

Both LMA and HSPY models satisfactorily fit the experimental data. Despite the LMA leading to an apparent better fit, note that there are several similar LMA-fits corresponding to very dissimilar size distributions. The one selected through our optimization criterion based on $\chi^2_{\rm HSPY}$ (see eq. (7)) which in our opinion is the most physically acceptable one.

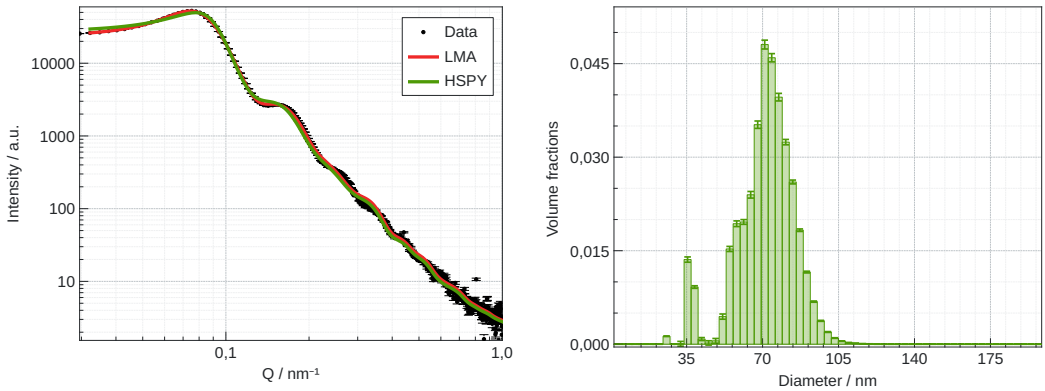


Scanning electron microscopy images of Sample 1 (courtesy of N. Feltin) for two distinct size scales. The two population modes at 70 nm and 30 nm respectively are clearly defined.

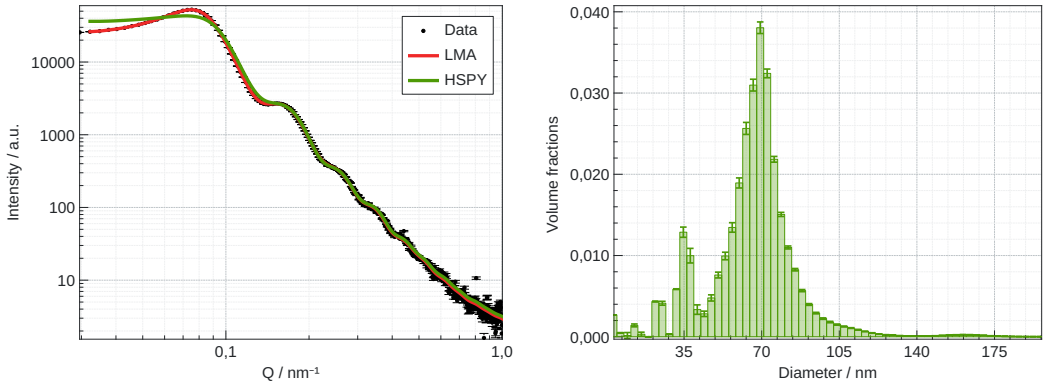
The fit in Fig. 4 features a disagreement between the models and the data in the mid Q range, i.e. from 0.2 to 0.4 nm^{-1}. One may find other LMA solutions (i.e. corresponding to a distinct $C^\star, \rho^\star$ pair) that satisfactorily describe this data range. In Fig. 5 we show the result we obtain if the optimization parameters are chosen to minimize the distance between the data and the LMA (instead of HSPY) model curve. While it clearly appears that LMA provides an excellent fit ($\chi^2 = 1.94$) to the experimental data, we do not depict it as the best solution due to the poor fit of the corresponding HSPY model intensity in the low Q region. In the particular case of Sample 1, both procedures (selection of best parameters based on HSPY and LMA respectively) lead to similar solutions (see Fig. 4 right and Fig. 5 right).

While in most cases one can find a good LMA fit to the data, the difficulty lies in

assessing whether one can trust the result. In general, and especially for very concentrated samples (e.g. dry powders), we have encountered excellent LMA fits associated to unphysical size distributions. This motivated us to choose HSPY for validating the solution, as in general one has no or little information on the sizes expected in the sample.



Sample 1: data fit (left) and volume fraction (right). The best fit (with respect to HSPY) is obtained for $C^\star = 0.96$ and $\rho^\star = 0.38$ ($\chi^2_{\mathrm{HSPY}} = 90$). The computed volume fraction features the two expected populations at 74 and 44 nm, respectively.



Sample 1: data fit (left) and volume fraction (right) corresponding to the best LMA solution. The best fit (with respect to LMA) is obtained for $C^\star = 1.04$ and $\rho^\star = 0.34$ ($\chi^2_{\mathrm{LMA}} = 1.94$). While this solution explains the mid Q range oscillations in the experimental data with both models, we discard it due to the poorer description of low Q data with the HSPY model, compared to the solution featured in Fig. 4.

While comparing the fits of the HSPY model in Figs. 4 ($\chi^2_{\mathrm{HSPY}} = 90$) and 5 ($\chi^2_{\mathrm{HSPY}} = 366$) we notice that the χ^2 criterion is more favorable to the HSPY model intensity in Fig. 4 as compared to the one in Fig. 5, although visually the contrary may be assumed. This is due to the fact the low Q points contribute with a larger weight to the χ^2 value, as compared to the higher Q points. Albeit the experimental uncertainties on the intensity values being well estimated, intensities obey the Poisson statistics and therefore the lower the Q value, the higher the signal to noise ratio. Therefore should the χ^2 criterion be employed, it favors the selection of a satisfactory fit in the low Q range (mainly signature of the particle interactions) than in the mid Q range (oscillations due to the particle sizes).

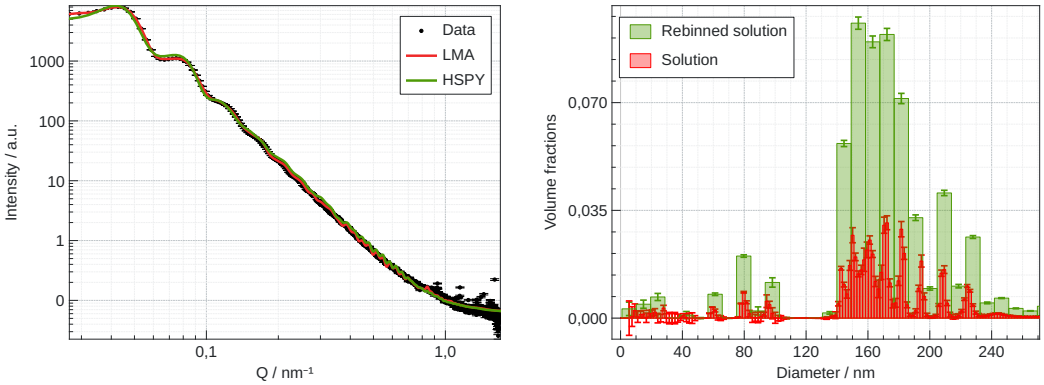
4.2. Sample 2: densely packed, dry powder

Sample 2 is a dry powder containing densely packed SiO_2 NP of stated 150 nm in diameter. SAXS measurements on Sample 2 were performed on a Xeuss device with a Pilatus 300k detector (Dectris) situated at 2622 mm from the sample and for a 60 minute exposure time. The data fit and solution are shown in Fig. 6. The sample is found to be very polydisperse with a main population mode at around 150 nm and several other lower size modes.

The very same SAXS data set (rebinned at high q values) was already interpreted within the LMA model solved by a Monte Carlo procedure (Bressler *et al.*, 2015). The authors find a main population at about 150 nm and a second mode at around 80 nm (Figure 7 in reference (Bressler *et al.*, 2015)) which is consistent with our results.

In the method proposed by reference (Bressler *et al.*, 2015) the parameter $\rho^\star$ is set manually. In our results, the optimized model parameter $\rho^\star = 0.65$ is close, but not identical to the value for the volume fraction expected for the close random packing of spheres (0.63) used in reference (Bressler *et al.*, 2015). The authors of (Bressler

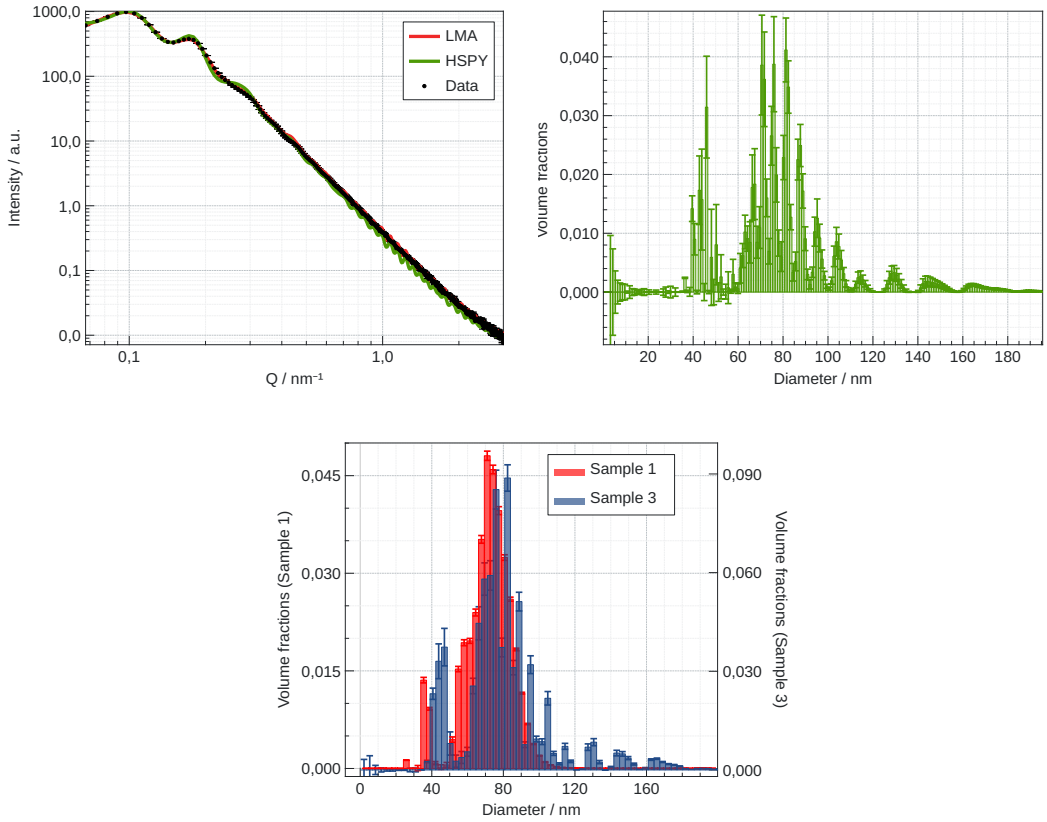
et al., 2015) point out that parameter $\rho^\star$, which they assimilate to the volume fraction of the dry powder, strongly affects the shape of the resulting distribution. We confirm this finding. The advantage of our method with respect to the one in reference (Bressler *et al.*, 2015) is that we are able to determine both the optimum model parameters and the size distribution thanks to the intrinsic advantage of EM over the Monte-Carlo techniques in terms of number of iterations to reach convergence.



Sample 2: data fit (left) and volume fraction (right). The best fit is obtained for $C^\star = 0.98$ and $\rho^\star = 0.65$, $\chi^2 = 760$, while the corresponding solution is noisy. To increase legibility we equally show its binned (5 by 5 points regrouping) version. The two population modes are consistent with the ones found by reference (Bressler *et al.*, 2015).

4.3. Sample 3: NP bipopulation in powder form

Sample 3 is the commercial product NexSil 85-40 (Nyacol, 2017) (same as Sample 1), furthermore dried to obtain a powder. The SAXS data acquisition was performed on the Nano-inXider in transmission geometry using Cu Kα radiation (λ = 1.54 Å) in high resolution mode. Scattering patterns were collected on a Pilatus3 (Dectris) detector, orthogonal to the beam and situated at approximately 1 meter from the sample, during 30 minute of exposure time.



Sample 3: data fit (left) and volume fraction (right). The best fit is obtained for $C^\star = 0.96$ and $\rho^\star = 0.71$, $\chi^2 = 1233$. The data fit is very good with respect to both LMA and HSPY model. The solution, while being extremely noisy, features the two expected populations at 45 nm and 85 nm. The bottom view contains the conveniently binned solution and its comparison to the size distribution obtained for Sample 1 (same as in Fig. 4) .

While the NP size distribution of Sample 3 is expected to be similar to the one of Sample 1, one can clearly see that the corresponding SAXS data are different. The challenge in treating Sample 3 is to assess whether we extract the same size information as for Sample 1.

To allow the comparison we have chosen a radial calculation grid whose step value is commensurate with the calculation step used with Sample 1 (Fig. 4 right). The data fit and the solution are shown in Fig. 7. As for the previous sample, the LMA and HSPY models are in reasonable agreement with the measured intensity. The value for

$C^\star = 0.97$ is close to 1 like the previous cases. We notice that the optimized $\rho^\star = 0.72$, which proves that in the case of powders the best LMA $\rho^\star$ parameter is not necessarily close to the value of the expected volume fraction for the closely packed spheres (0.63 according to (Song *et al.*, 2008)). This furthermore implies that the previously cited theoretical value should not be used as an upper bound of the search grid for $\rho^\star$ parameter optimization.

The solution in in Fig. 7 top right is noisy, both in the sense of large calculation uncertainties and in the oscillating behavior. Furthermore, this leads to high frequency oscillations of model intensities in the Porod (high Q) region. These features could in principle be suppressed by a noise reduction scheme such as regularization, which lies beyond the scope of our work. Instead, to reduce noise and facilitate the comparison with the results obtained for Sample 1, we binned the solution 3 by 3 points. The result of the comparison is displayed at the bottom of Fig. 7.

The two population modes are identified at 80 and 45 nm, respectively, where the mean values have been estimated as $\sum_k \rho_k \cdot R_k / \sum_k \rho_k$ on the corresponding ranges. The population modes depicted in Sample 3 are close but not identical to the modes of Sample 1. This may come from numerical artifacts, model limitations resulting from the use of the LMA model or sample modifications during the drying process.

One may expect that the size distribution of Sample 3 is less noisy than the one of Sample 1 since the signal to noise ratio in the experimental data is more favorable for the former. However, the application of EM-LMA leads to the contrary. This can be understood by noting that LMA is more likely to be a pertinent model for Sample 1 than it is for Sample 3 (the denser the sample, the less likely for interactions to occur locally). One could regularize the solution of Sample 3 as to render it smoother but this would involve extra calculation parameters and is therefore beyond the scope of the present work.

4.4. Discussion

While the results presented in this section are satisfactory in terms of model *versus* experimental data fits, we emphasize the fact that LMA, under whatever implementation, is not expected to be an all-time solution on all concentrated NP samples, due to its intrinsic simplifications. Having some *a priori* knowledge on the sample like the true volume fraction or the polydispersity can help, not only to initialize the algorithm such as to spare calculation time, but to validate whether the result is physically acceptable. One should keep in mind that a good data fit is no guarantee the size distribution actually describes the real content of the sample. Indeed, if, after having explored the entire space of model parameters $\{C^\star, \rho^\star\}$, no reasonable fit was found, it implies the LMA model is not a good enough approximation in order to describe the interactions in the sample. Nonetheless, a good data fit with respect to both the full HSPY model and its simplifying limit LMA is a strong factor in favor of the validation of the solution.

We believe the $\chi^2 \leq 1$ condition is sufficient, but not necessary, to assess the data fit as satisfactory. The numerical values of χ^2 at convergence are strongly related to the signal to noise ratio in the data. Dense samples have strong scattering power and therefore the signal to noise ratio in the corresponding SAXS data is elevated. Under these conditions data fits can be satisfactory in spite the high χ^2 values.

5. Conclusion

The original LMA implementation proposed in this work allowed us to successfully solve the NP sizing problem on two NP powders and a concentrated dispersion measured with the SAXS technique on a laboratory instrument. We took benefit from the computational efficiency of the *EM* algorithm to perform a robust brute force exploration of the LMA model parameters and hence make sure of the optimality of the final result. Moreover, our strategy relies on cross-validating the LMA and the HSPY

models such as to discard ill (un-physical) solutions associated with good data fits and wrong LMA parameters. By cross-validation we mean the comparison between the best LMA fit and the HSPY intensity associated to the corresponding LMA solution.

The advantage of our EM-LMA implementation with respect to alternative LMA implementations is the enhanced usability. Firstly, there is no requirement for user-adjustable values : the LMA model parameters are optimized automatically and there are no other calculation parameters involved. Secondly, EM-LMA is intrinsically more reliable than any other general curve fitting scheme as it makes no assumption on the size distribution (e.g. as in shape or mean radius). Furthermore the implementation is general and can be used in the very same form on both diluted or dense NP samples. Altogether we feature a general, robust and easy to use method to extract NP sizes from SAS data with no assumptions on the sample content.

The main focus of this work is to feature a robust and highly efficient numerical implementation of the LMA model with an EM algorithm. Contrary to alternative implementations of LMA, ours is parameter-free in the sense that an automatic search is performed in order to find the best model parameters with respect to the fitted data. The accuracy of the fit is measured using the more complete HSPY model. This automatic parameter optimization is feasible due to the low computational cost of EM as compared to Monte-Carlo or variational methods.

Should the LMA model fail to describe the experimental data, the full HSPY model must be inverted instead. Its non-linear nature does not make HSPY a good candidate for EM-based solving. A substantial validation of this model applied for particle sizing in dense samples and the description of the associated inversion procedure will the subject of a forthcoming publication.

6. APPENDIX

6.1. Expectation Maximization

The EM optimization method is a well-known fix point algorithm which aims at maximizing the likelihood of obtaining the (possibly noisy) data y^δ given a set of parameters x (see (Natterer & Wübbeling, 2001) p. 45 and p. 118). In the case of data ruled by Poisson statistics and by assuming a linear model $y = \mathbf{H}x$ with positive entries for both the data y and the matrix of the model $\mathbf{H}$, the likelihood function reads

$$p(y^\delta|x) = \prod_{i=1}^{m} \frac{e^{(\mathbf{H}x)_i}}{y_i^\delta!} (\mathbf{H}x)_i^{y_i^\delta} \tag{8}$$

The maximization is strictly equivalent to minimizing the log-likelihood defined by

$$L(y^\delta|x) = -\log(p(y^\delta|x)). \tag{9}$$

Writing the optimality condition for the min argument of eq. 9 we obtain the classical EM algorithm as a fixed-point iteration to solve for the solution (see for instance (Natterer & Wübbeling, 2001)).

By computing the Hessian of eq. 9 it is possible to show that the log-likelihood function is strictly convex for the likelihood function given in (8), when the entries for y^δ and $\mathbf{H}$ are positive. This means that under these conditions the iterative method is formally guaranteed to converge. In other words, the EM algorithm yields the only solution to the inverse problem for a given right hand side y^δ and therefore the solution is not sensitive to the initial guess.

The EM optimization scheme belongs to the class of deterministic methods (iteration n is conditioned by the result of the previous iteration $n-1$) and therefore involves significantly less iterations than Monte Carlo approaches in order to reach a given goodness of fit.

6.2. The partial structure factor

The partial structure factor is related to the total correlation functions h_{kl} by the equation (see (Guinier & Fournet, 1955) p. 60 to 82) :

$$S_{kl}(Q) = \delta_{kl} + \sqrt{n_k n_l}\frac{1}{V}\int_0^\infty 4\pi r^2 h_{kl}(r)\frac{\sin(Qr)}{Qr}\,dr \tag{10}$$

where in the dilute limit (eq. 1) the total correlation functions $h_{kl} = 0$. The functions h_{kl} are related to the direct correlation functions c_{kl} by the Ornstein-Zernike equation (see 3.2 in (Torquato, 2002))

$$h_{kl}(r) = c_{kl}(r) + \sum_{p=1}^{n}\frac{n_p}{V}\int_{\mathbf{s}\in\mathbb{R}^3} c_{kp}(|\mathbf{s}|)h_{pl}(|\mathbf{s}-\mathbf{r}|)\,d\mathbf{s} \tag{11}$$

A closure relation is needed to determine c_{kl}, h_{kl} and consequently S_{kl}. This relation links the direct correlation function to the interaction potential ϕ_{kl} between two particles of kind k and l.

Equation (11) is solved in the Fourier domain as it features convolution products. In general this solution is only accessible numerically. However, for some couples $\{c_{kl}, \phi_{kl}\}$, the partial structure factor (10) can be calculated analytically, making much cheaper the computational cost. This is in particular the case of hard spheres interaction potentials with the closure relation of Percus-Yevick (HSPY).

According to (Kinning & Thomas, 1984) the partial structure factor within the LMA model is :

$$S(Q, \rho^\star, C^\star) = \frac{1}{1 + 24\,\rho^\star\dfrac{G(A,\rho^\star)}{A}} \tag{12}$$

where

$$A = 2 \cdot Q \cdot C^\star \cdot R \tag{13}$$

and

$$\begin{aligned} G(A) = & \quad \frac{\alpha}{A^2}(- & A\cos(A) + \sin(A)) \\ + & \quad \frac{\beta}{A^3}(- & 2 + 2\,A\sin(A) + (2 - A^2)\cos(A)) \\ + & \quad \frac{\gamma}{A^5}(- & A^4\cos(A) + 4((3\,A^2 - 6)\cos(A) \\ & \quad + & (A^3 - 6\,A)\sin(A) + 6)) \end{aligned} \tag{14}$$

with

$$\begin{aligned} \alpha &= (1 + 2\rho^\star)^2/(1 - \rho^\star)^4, \\ \beta &= -6\,\rho^\star(1 + 0.5\rho^\star)^2/(1 - \rho^\star)^4, \\ \gamma &= \frac{\rho^\star}{2}\alpha. \end{aligned}$$

6.3. The HSPY model

An interacting particle model is determined by the form of the interaction potential and by the choice of the closure relation. In the case of HSPY these are the hard spheres potential with the Percus-Yevick closure, respectively. The interaction potential between particles k and l is:

$$\phi_{kl}(r) = \begin{cases} \infty, & 0 \le r \le R_k + R_l, \\ 0, & r > R_k + R_l. \end{cases} \tag{15}$$

i.e. a contact interaction involving non-penetrating spheres. In particular, $c_{kl}(r) = 0$ whenever $r > R_k + R_l$. The Percus-Yevick closure relation is well-suited for short ranged potentials such as the one above and reads:

$$c_{kl}(r) = (1 - e^{-\frac{\phi_{kl}(r)}{\kappa T}}) \cdot (h_{kl}(r) + 1) \tag{16}$$

In this framework, a fully analytical expression of each partial structure factor (10) is given in reference (Vrij, 1979). However, computing each $S_{kl}(Q)$ individually is expensive and not numerically recommended in the context of iterative inversion methods.

One prefers directly using the analytical expression for the normalized intensity that also can found in (Vrij, 1979).

We reproduce the main steps of this computation below. Let us define some notations following those of (Vrij, 1979) :

$$d_i = 2 \cdot R_i,$$

$$X_i = Q \cdot R_i,$$

$$\xi_\nu = \frac{\pi}{6} \sum_{i=1}^{n} n_i d_i^\nu,$$

$$\langle Y \rangle = \frac{\pi}{6} \sum_{i=1}^{n} n_i Y_i.$$

where n_i is the number density of the i species of radius R_i. The steps for the computation of the direct model intensity for a given scattering vector Q are the following:

$$
\begin{aligned}
\Psi_i &= \frac{\sin(X_i)}{X_i} \\
\Phi_i &= 3\frac{\Psi_i - \cos(X_i)}{X_i^2}, \\
M_i &= \frac{\pi}{6} \cdot \frac{d_i^3 \cdot \Phi_i}{1-\xi_3} \\
N_i &= \frac{\pi}{6} \cdot \frac{d_i^2}{1-\xi_3} \left(3 \cdot \Psi_i - \mathbf{i} X_i \cdot \Phi_i + 3/(1-\xi_3) \cdot \xi_2 \cdot d_i \cdot \Phi_i\right), \\
F_{11} &= 1 - \xi_3 + \langle d^3 \cdot e^{\mathbf{i}X} \cdot \Phi \rangle, \\
F_{12} &= \langle d^4 \cdot e^{\mathbf{i}X} \cdot \Phi \rangle, \\
F_{21} &= \frac{1}{2}(1-\xi_3)\mathbf{i}Q - 3 \cdot \xi_2 + 3\langle d^2 \cdot e^{\mathbf{i}X} \cdot \Psi \rangle, \\
F_{22} &= 1 - \xi_3 + 3\langle d^3 \cdot e^{\mathbf{i}X} \cdot \Psi \rangle, \\
T_1 &= F_{11} \cdot F_{22} - F_{12} \cdot F_{21}, \\
T_2 &= F_{21}\langle d \cdot f \cdot e^{\mathbf{i}X} \rangle - F_{22}\langle f \cdot e^{\mathbf{i}X} \rangle, \\
T_3 &= F_{12}\langle f \cdot e^{\mathbf{i}X} \rangle - F_{11}\langle d \cdot f \cdot e^{\mathbf{i}X} \rangle, \\
\Delta &= \frac{|T_1|^2}{(1-\xi_3)^4}, \\
D_f &= -\frac{6}{\pi(1-\xi_3)^4} \cdot \Big(\langle f^2 \rangle \cdot |T_1|^2 + \langle d^6 \cdot \Phi^2 \rangle \cdot |T_2|^2 + 9\langle d^4 \cdot \Psi^2 \rangle \cdot |T_3|^2 \\
&\quad + \langle f \cdot d^3 \cdot \Phi \rangle \cdot 2\Re(T_1 T_2^\star) + 3\langle f \cdot d^2 \cdot \Psi \rangle \cdot 2\Re(T_1 T_3^\star) \\
&\quad + 3\langle d^5 \cdot \Phi \cdot \Psi \rangle \cdot 2\Re(T_2 T_3^\star) \Big).
\end{aligned}
$$

Finally, we compute

$$
I_{\mathrm{HSPY}} = -\frac{D_f}{\Delta}. \tag{17}
$$

Since the dependency of I with respect to ρ is non-linear, an accurate knowledge of the scattering length density is mandatory in order to compute the HSPY model accurately. Note that the LMA model is not equivalent to taking the diagonal of the structure factor matrix $(S_{kl})_{k,l}$ relative to the HSPY model (Vrij, 1978): the latter takes explicitly into account the unknown size distribution ρ while it is not the case

for the structure factor estimation of the LMA model.

The calculation of the HSPY model intensity corresponding to the LMA solution (instead of the real size distribution) of the scattering problem in the case of powders requires us to introduce two additional parameters $\{\alpha, \beta\}$ which we develop below.

SAS data acquired on powders is generally not fully normalized : the lack of accuracy in the determination of the sample thickness in the direction of the beam leads to a experimental intensities being normalized to absolute units in the limit of an arbitrary, Q independent, scaling factor (Spalla *et al.*, 2003). In the case where the measured intensity is not available in absolute units, there exists a multiplicative constant α between the model and the measure such that

$$I_{\text{measure}} \equiv I^{\text{real}}_{\text{HSPY}} = \alpha \cdot I_{\text{HSPY}}(\rho).$$

If the model is linear (e.g. LMA), we have $\alpha \cdot \mathbf{H} \cdot \rho = \mathbf{H} \cdot (\alpha\rho) = \mathbf{H} \cdot \tilde{\rho}$ where ρ is the true (physical) solution and $\tilde{\rho}$ is the actual solution computed by the inversion algorithm. It means that the solution is only obtained up to a $1/\alpha$–multiplicative factor. The cause is that the inversion algorithm is given $\mathbf{H}$, not $\alpha \cdot \mathbf{H}$.

In the non-linear case (HSPY), this constant cannot be merged with the solution $\tilde{\rho}$. It means that we must introduce a second constant β such that

$$I^{\text{real}}_{\text{HSPY}} = \alpha \cdot I_{\text{HSPY}}(\beta \cdot \tilde{\rho})$$

if we wish to feed the HSPY model with the LMA solution $\tilde{\rho}$. If the problem were linear, β would have identified to $1/\alpha$.

In practice $\{\alpha, \beta\}$ are optimized on a linear grid. We stress upon the fact that the optimization of α is only needed when one lacks absolute intensity units (not fully normalized data). On the other hand, the optimization of β is only needed for the calculation of the HSPY model. In the case where the experimental data is fully normalized and the LMA model parameters are selected by minimizing the distance between the

LMA model curve and the data, there is no need for the $\{\alpha, \beta\}$ optimization step.

This work was funded by the FUI project SAXSize.

We thank Nicolas Feltin for providing the SEM images and their interpretation. We thank Gemma Newby, Blandine Lantz and Bertrand Faure for having carefully read the manuscript and provided useful comments. All the figures in this manuscript have been produced with XSACT (X-ray Scattering Analysis and Computation Tool), a proprietary software from Xenocs (Xenocs, 2019).

/* We omit the dependency in Q. */
Input: ρ^0, I_{ref}, σ, ε_{EM}, $\{C^\star\} = \{C_1^\star, \ldots, C_{n_C}^\star\}$, $\{\rho^\star\} = \{\rho_1^\star, \ldots, \rho_{n_\rho}^\star\}$, $\{\alpha^\star\} = \{\alpha_1^\star, \ldots, \alpha_{n_\alpha}^\star\}$, $\{\rho_{\min}, \rho_{\max}\}$
$\chi^2_{\text{opt}} \leftarrow 10^{30}$
for $C^\star \in \{C^\star\}$ **do**
  for $\rho^\star \in \{\rho^\star\}$ **do**
    $\mathbf{H}_{\text{LMA}} \leftarrow \mathbf{H}_{\text{dil}} : \mathbf{S}(C^\star, \rho^\star)$
    $\rho, I_{\text{LMA}} \leftarrow \text{EM}(I_{\text{ref}}, \rho^0, \varepsilon_{\text{EM}}, \mathbf{H}_{\text{LMA}})$
    $\beta_1^\star = \dfrac{\rho_{\min}}{\sum \rho_i}$, $\beta_{n_\beta}^\star = \dfrac{\rho_{\max}}{\sum \rho_i}$
    $\{\beta^\star\} = \{\beta_1^\star, \ldots, \beta_{n_\beta}^\star\}$
    $\chi_t^2 = 10^{30}$
    for $\beta^\star \in \{\beta^\star\}$ **do**
      $I \leftarrow I_{\text{HS}}(\beta^\star \cdot \rho)$
      for $\alpha^\star \in \{\alpha^\star\}$ **do**
        $\chi_0^2 \leftarrow \chi^2(I_{\text{ref}}, \alpha^\star \cdot I, \sigma)$
        if $\chi_0^2 < \chi_t^2$ **then**
          $\chi_t^2 \leftarrow \chi_0^2$, $\alpha_t^\star \leftarrow \alpha^\star$, $\beta_t^\star \leftarrow \beta^\star$, $I_t \leftarrow I$
        end if
      end for
    end for
    if $\chi_t^2 < \chi^2_{\text{opt}}$ **then**
      $\rho_{\text{opt}} \leftarrow \rho$, $I_{\text{HS,opt}} \leftarrow I_t$, $\chi^2_{\text{opt}} \leftarrow \chi_t^2$, $C^\star_{\text{opt}} \leftarrow C_t^\star$, $\rho^\star_{\text{opt}} \leftarrow \rho_t^\star$, $\alpha^\star_{\text{opt}} \leftarrow \alpha_t^\star$, $\beta^\star_{\text{opt}} \leftarrow \beta_t^\star$
    end if
  end for
end for
return ρ_{opt}, $I_{\text{LMA,opt}}$, $C^\star_{\text{opt}}$, $\rho^\star_{\text{opt}}$, χ^2_{opt}, $\alpha^\star_{\text{opt}}$, $\beta^\star_{\text{opt}}$

EM algorithm for the inversion of the SAXS problem – Fitting with respect to the full HSPY model

References

Benvenuto, F., Haddar, H. & Lantz, B. (2016). *SIAM J. Appl. Math.* **76**(1), 276–292.

Breßler, I., Kohlbrecher, J. & Thünemann, A. F. (2015). *Journal of Applied Crystallography*, **48**(5), 1587–1598.

Bressler, I., Pauw, B. R. & Thünemann, A. F. (2015). *Journal of Applied Crystallography*, **48**, 962–969.

Brunner-Popela, J. & Glatter, O. (1997). *Journal of Applied Crystallography*, **30**(4), 431–442.

Brunner-Popela, J., Mittelbach, R., Strey, R., Schubert, K.-V., Kaler, E. W. & Glatter, O. (1999). *The Journal of Chemical Physics*, **110**(21), 10623–10632.

Egami, T. & Billinge, S. J. L. (2003). *Underneath the Bragg Peaks, Structural Analysis of Complex Materials*, vol. 7 of *Pergamon Material Series*. Pergamon.

Ehmann, H. M. A., Spirk, S., Doliška, A., Mohan, T., Gössler, W., Ribitsch, V., Sfiligoj-Smole, M. & Stana-Kleinschek, K. (2013). *Langmuir*, **29**(11), 3740–3748.

Feigin, L. A. & Svergun, D. I. (1987). *Structure Analysis by Small-Angle X-ray and Neutron Scattering*. Springer New York.

Feltin, N., (2018*a*). Première comparaison inter-techniques et inter-laboratoires française pour la caractérisation de la taille de nanoobjets. `http://www.pcipresse.fr/spectra-analyse-par-article/1055-n-321-pages-29-a-36.html`.

Feltin, N., (2018*b*). private communication.

Glatter, O. (1977). *Journal of Applied Crystallography*, **10**, 415–421.

Guinier, A. & Fournet, G. (1955). *Small-Angle Scattering of X-rays*. Structure of Matter Series. J. Wiley and Sons Inc.

Kinning, D. J. & Thomas, E. L. (1984). *Macromolecules*, **17**, 1712–1718.

Kremser, G., Rath, T., Kunert, B., Edler, M., Fritz-Popovski, G., Resel, R., Letofsky-Papst, I., Grogger, W. & Trimmel, G. (2012). *Journal of Colloid and Interface Science*, **369**(1), 154 – 159.

Natterer, F. & Wübbeling, F. (2001). *Mathematical Methods in Image Reconstruction*. Society for Industrial and Applied Mathematics.

Nyacol, (2017). Nexsil 85-40. `https://www.nyacol.com/wp-content/uploads/2015/04/NexSil-85-40-Data-Sheet.pdf`.

Pauw, B. R., Pedersen, J. S., Tardif, S., Takata, M. & Iversen, B. B. (2013). *Journal of Applied Crystallography*, **46**(2), 365–371.

Pauw, B. R., Smith, A. J., Snow, T., Terrill, N. J. & Thünemann, A. F. (2017). *Journal of Applied Crystallography*, **50**(6), 1800–1811.

Pedersen, J. S. (1994). *Journal of Applied Crystallography*, **27**, 595–608.

Pedersen, J. S. (1997). *Adv. Colloid Interface Sci.* **70**, 171–210.

Potocnik, J., (2011). Commission recommendation of 18 october 2011 on the definition of a nanomaterial. `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32011H0696`.

Rasmussen, K., González, M., Kearns, P., Sintes, J. R., Rossi, F. & Sayre, P. (2016). *Regulatory Toxicology and Pharmacology*, **74**, 147 – 160.

Rieker, T., Hanprasopwattana, A., Datye, A. & Hubbard, P. (1999). *Langmuir*, **15**(2), 638–641.

Song, C., Wang, P. & Makse, H. A. (2008). *Nature*, **453**, 629 EP –.

Spalla, O., Lyonnard, S. & Testard, F. (2003). *Journal of Applied Crystallography*, **36**(2), 338–347.

Torquato, S. (2002). *Random Heterogeneous Materials*. Interdisciplinary Applied Mathematics. Springer-Verlag New York.

Vrij, A. (1978). *The Journal of Chemical Physics*, **69**, 1742–1747.

Vrij, A. (1979). *The Journal of Chemical Physics*, **71**, 3267–3270.

Weyerich, B., Brunner-Popela, J. & Glatter, O. (1999). *Journal of Applied Crystallography*, **32**(2), 197–209.

Xenocs, (2019). XSACT: X-ray Scattering Analysis and Calculation Tool. `www.xenocs.com/products/software`. SAXS & WAXS data analysis software – Version 1.0.

Synopsis

This paper presents a robust method for the resolution of SAS problems featuring a structure factor. It is based on the Local Monodisperse Approximation and the Expectation-Maximization algorithm.